B3

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 April 2001 (26.04.2001)

PCT

(10) International Publication Number
WO 01/29217 A2

(51) International Patent Classification[7]: C12N 15/12, 15/16, C07K 14/575, 14/705, 14/47, G01N 33/68, C12Q 1/68, C07K 16/18, 16/26, 16/28, C12N 5/10, A61K 31/7088, 38/17, 38/22, 39/395, 48/00

(21) International Application Number: PCT/US00/28474

(22) International Filing Date: 13 October 2000 (13.10.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/159,805	15 October 1999 (15.10.1999)	US
60/159,992	18 October 1999 (18.10.1999)	US
60/160,952	22 October 1999 (22.10.1999)	US
Not furnished	12 October 2000 (12.10.2000)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/159,805 (CIP)
Filed on	15 October 1999 (15.10.1999)
US	60/159,992 (CIP)
Filed on	18 October 1999 (18.10.1999)
US	60/160,952 (CIP)
Filed on	22 October 1999 (22.10.1999)
US	Not furnished (CIP)
Filed on	12 October 2000 (12.10.2000)

(71) Applicant *(for all designated States except US)*: **CURAGEN CORPORATION** [US/US]; 555 Long Wharf Drive, 11th floor, New Haven, CT 06511 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **PRAYAGA, Sudhirdas, K.** [IN/US]; 140 Mill Street, Apt. 18-136, East Haven, CT 06512 (US). **TAUPIER, Raymond, J., Jr.** [US/US]; 407 Holmes Street, East Haven, CT 06512 (US). **BANDARU, Raj** [IN/US]; 3C 107 Cottage Street, New Haven, CT 06511 (US).

(74) Agent: **ELRIFI, Ivor, R.**; Mintz, Levin, Cohn, Ferris, Glovsky, and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

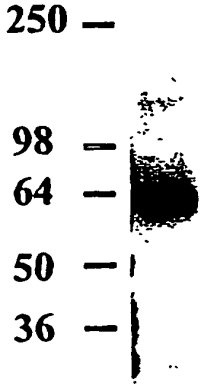
Published:

— *Without international search report and to be republished upon receipt of that report.*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES ENCODING SAME

(57) Abstract: Disclosed herein are novel human nucleic acid sequences. Also disclosed are polypeptides encoded by these nucleic acid sequences, and antibodies which immunospecifically-bind to the polypeptide, as well as derivatives, variants, mutants, or fragments of the aforementioned polypeptide, polynucleotide, or antibody. The invention further discloses therapeutic, diagnostic and research methods for diagnosis, treatment, and prevention of disorders involving these novel nucleic acids and proteins.



WO 01/29217 A2

In a further aspect, the invention includes a method of producing a polypeptide by culturing a cell that includes a NOV nucleic acid, under conditions allowing for expression of the NOV polypeptide encoded by the DNA. If desired, the NOV polypeptide can then be recovered.

In another aspect, the invention includes a method of detecting the presence of a NOV polypeptide in a sample. In the method, a sample is contacted with a compound that selectively binds to the polypeptide under conditions allowing for formation of a complex between the polypeptide and the compound. The complex is detected, if present, thereby identifying the NOV polypeptide within the sample.

Also included in the invention is a method of detecting the presence of a NOV nucleic acid molecule in a sample by contacting the sample with a NOV nucleic acid probe or primer, and detecting whether the nucleic acid probe or primer bound to a NOV nucleic acid molecule in the sample.

In a further aspect, the invention provides a method for modulating the activity of a NOV polypeptide by contacting a cell sample that includes the NOV polypeptide with a compound that binds to the NOV polypeptide in an amount sufficient to modulate the activity of said polypeptide. The compound can be, e.g., a small molecule, such as a nucleic acid, peptide, polypeptide, peptidomimetic, carbohydrate, lipid or other organic (carbon containing) or inorganic molecule, as further described herein.

Also within the scope of the invention is the use of a therapeutic in the manufacture of a medicament for treating or preventing disorders or syndromes outlined in the preferred embodiment below. The therapeutic can be, e.g., a NOV nucleic acid, a NOV polypeptide, or a NOV-specific antibody, or biologically-active derivatives or fragments thereof.

In the preferred embodiments, the invention further includes methods for screening for a modulator of disorders or syndromes including, e.g., those involving development, differentiation, and activation of thymic immune cells; in pathologies related to spermatogenesis and male infertility; diagnosis of several human neoplasias; in diseases or pathologies of cells in blood circulation such as red blood cells and platelets; neurological, cardiac and vascular pathologies; rheumatoid arthritis; congenital muscular dystrophies; various muscle disorders; fixed deformities (arthrogryposis); small cell lung cancer NCI-H23; prostate cancer; and abnormal white matter. The method includes contacting a test compound with a NOV polypeptide and determining if the test compound binds to said NOV polypeptide. Binding of the test compound to the NOV polypeptide indicates the test

advantages of the invention will be apparent from the following detailed description and claims.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1. Western blot of a NOV2 polypeptide secreted by 293 cells.

Figure 2. Western blot of a NOV3 polypeptide secreted by 293 cells.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides novel nucleotides and polypeptides encoded thereby. Included in the invention are the novel nucleic acid sequences and their polypeptides. The sequences are collectively referred to as "NOV nucleic acids" or "NOV polynucleotides" and the corresponding encoded polypeptides are referred to as "NOV polypeptides" or "NOV proteins." Unless indicated otherwise, "NOV" is meant to refer to any of the novel sequences disclosed herein. Table 11 provides a summary of the NOV nucleic acids and their encoded polypeptides.

NOV nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOV nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins.

For example, NOV1 is homologous to members of the thymosin beta 10 family of proteins. As a result, NOV1 has various marker utilities as described herein. Also, NOV1 has efficacy in treatment of conditions involving development, differentiation, and activation of thymic immune cells; in pathologies related to spermatogenesis and male infertility; diagnosis of several human neoplasias; in diseases or pathologies of cells in blood circulation such as red blood cells and platelets; and detection of small cell lung cancer. NOV2 is homologous to members of the ephrin A receptor family. As a result, NOV2 has various marker utilities as described herein. NOV2 has efficacy in the treatment of conditions involving neurological, cardiac and vascular pathologies. NOV2 also has utility in the detection of prostate cancer. NOV3 is homologous to members of the proteoglycan family. As a result, NOV3 has various marker utilities as described herein. NOV3 also has efficacy in the treatment of conditions involving rheumatoid arthritis; congenital muscular dystrophies; various muscle disorders; fixed deformities (arthrogryposis); and abnormal white matter. Additional utilities for NOV nucleic acids and polypeptides according to the invention as also discussed herein.

The thymosin-betas comprise a family of structurally related, highly conserved acidic polypeptides that sequester actin and regulate actin dynamics within cells. During embryogenesis the control of actin polymerization is essential in processes such as cell migration, angiogenesis and neurogenesis. Direct visualization and quantitation of actin filaments has shown that thymosin-betas, like agonists, induced actin depolymerization at the apical membrane where exocytosis occurs (Muallem S, Kwiatkowska K, Xu X, Yin HL, J Cell Biol 1995 Feb;128(4):589-98). Thymosin-beta-10 is widely distributed in mammalian tissues including the nervous system, and the presence of this transcript in different regions of the rat forebrain, including hippocampus, neocortex and several brain nuclei, provides evidence for the participation of thymosin-beta-10 in the control of the actin dynamics that takes place in neurons. Thymosin-beta-10 is expressed at relatively high levels in embryonic and developing tissues (Hall AK *Cell Mol Biol Res* 1995;41(3):167-80), and given that it is involved in the inhibition of actin polymerization, the thymosin-beta-10 protein-like proteins can play an important role in early development.

mRNA species of similar molecular weights encoding thymosin beta-10 are found in most tissues of the rat; however, Lin and Morrison-Bogorad (J Biol Chem. 1991 Dec 5;266(34):23347-53) identified an additional thymosin-beta-10 mRNA of higher molecular weight in the testis of sexually mature rats. The latter mRNA differs from the ubiquitous form only in its 5-prime untranslated region, beginning 14 nucleotides upstream of the translation initiation codon. This finding, together with primer extension experiments, suggested that the two mRNA types are transcribed from the same gene through a combination of differential promoter utilization and alternative splicing. Both mRNAs were present in pachytene spermatocytes; only the testis-specific mRNA was detected in postmeiotic haploid spermatids. Immunohistochemical analysis showed that the protein was present in differentiating spermatids, suggesting that the testis-specific thymosin-beta-10 mRNA is translated in haploid male germ cells. Immunoblot analysis using specific antibodies showed that the thymosin-beta-10 protein synthesized in adult testis was identical in size to that synthesized in brain.

Thymosin-beta-10-like proteins also influence several properties of lymphocytes including cyclic nucleotide levels, migration inhibitory factor production, T-dependent antibody production, as well as the expression of various cell surface maturation/differentiation markers (Bodey B, Bodey B Jr, Siegel SE, Kaiser HE Int J Immunopharmacol 2000 Apr;22(4):261-73). These and other observations suggest that thymosin beta-10 (a) plays a significant and possibly obligatory role in cellular processes controlling apoptosis possibly by acting as an actin-mediated tumor suppressor, (b) functions

tissue types. Thymosin-beta-10 mRNA is overexpressed in a variety of tumors and tumor cell lines. Expression levels of thymosin-beta-10 like nucleic acids such as NOV1 are also useful in distinguishing T cell types given that expression of various cell surface/differentiation markers is influenced by thymosin-beta-10 like proteins such as a NOV1 polypeptide. A NOV1 nucleic acid has enhanced expression in certain cancer cell lines, especially non-small cell lung cancer NCI-H23, but not in cell lines from the corresponding normal tissue; therefore, NOV1 nucleic acids are useful as a cancer specific marker in such tissues (Example 1).

Given that thymosin-beta-10 related proteins can sequester actin and regulate actin dynamics within cells, proteins related to the NOV1 polypeptide are useful in screens for test compounds that can modulate actin polymerization or the formation or stability of actin-thymosin beta-10 complexes. Finally, since thymic hormones strengthen the effects of immunomodulators in immunodeficiencies, autoimmune diseases, and neoplastic malignancies, NOV1 related proteins can be used in combined chemo-immunotherapeutical anti-cancer treatments.

Table 1.
A representative cDNA sequence encoding the thymosin-beta-10-like protein according to the invention

Putative untranslated regions are underlined. The start and stop codons are in bold type.

GCCAGCAGGAGTGCCATGGTGAGAGGCACTGGCAGGGAATGCTAGGATTGTTTTAAGAAA**ATG**GCAGACAAACCAGACATAGGG
GAAATCGCCAGCTTCAATAAGGCCAAGCTGAAGAAAACAGAGATGCAGGAGAACACCCTGCTGACCAAAGAGGCCATTGAGCAG
GAGAAGCGGGTGAAATTTCCTAAGAGCCTGGAGGATTCCCTACCCCTGTCATCTTCGAGACCCCAG**TAG**TAATGTGGAGGAAGA
ATCACCACAAGATGGACACAAGCCACAAACTGTGACGTGAACCTGGGCACTCCGTGCTGATGCCACCAGCCTGAGGGTCCCTAT
GGGTCCAATCAGACTGCCAAATTCTCTGGTTTGCCCTGGGATATTATAGAAAATTATTTGCGTGAATAATGAAAACACAGCTCA
TGGCAAAAAA (SEQ ID NO:1)

A representative amino acid sequence of the thymosin-beta-10-like protein according to the invention

MADKPDIGEIASFNKAKLKKTEMQENTLLTKEAIEQEKRVKFPKSLEDSLPLSSSRPQ (SEQ ID NO:2)

Table 2.
Comparison between a NOV1 polypeptide and thymosin beta-10 from human

```
>gb|AAA36746.1| (M92383) thymosin beta-10 [Homo sapiens]
          Length = 49

 Score = 84.5 bits (192), Expect = 3e-16
 Identities = 34/40 (85%), Positives = 36/40 (90%), Gaps = 1/40 (2%)

NOV1 : 1  MADKPDIGEIASFNKAKLKKTEMQE-NTLLTKEAIEQEKR 39
          ||||||+|||||||+|||||||||| || ||| ||| ||||||
```

The representative ORF includes a 992 amino acid polypeptide (SEQ ID NO:5). The encoded polypeptide has a high degree of homology (approximately 95 percent identity) with mouse ephrin type-A receptor 8 precursor (Table 6) (SWISSPROT ACC: O09127, 956 out of 1005 residues). The NOV2 polypeptide also has an even higher degree of homology (100 percent identity) to a human eph- and elk-related kinase known as ephrin receptor EphA8 (Table 6A, partial sequence disclosed in Chan et al. (1991) Oncogene 6 1057-1061; the full length human ephrin receptor EphA8 full length sequence was deposited in Genbank September 14, 2000 as accession number NP_065387.1.) A multiple alignment with similar proteins showed comparable degrees of similarity to ephrin receptors from mouse (EPA8_mouse), human (EPA5_human), and chicken (EPA5_chick) (Table 7). In the predicted extracellular domain, a cysteine-rich region and tandem fibronectin type III repeats are present while a catalytic domain is present in the intracellular domain. These features are consistent with other members of the Eph family. Based on its relatedness, the NOV2 protein is a member of the ephrin type-A receptor tyrosine-protein kinase family.

The Eph receptors constitute the largest known family of receptor protein tyrosine kinases. They have been implicated in mediating developmental events, particularly in the nervous system. Receptors in the Eph subfamily typically have a single kinase domain and an extracellular region containing a Cys-rich domain and two fibronectin type III repeats. These receptors play important roles along with their ligands, called ephrins, in neural development, angiogenesis, and vascular network assembly. (Choi S, Jeong J, Kim T, Park S., Mol. Cells 9(4):440-45 (1999)).

The ephrin type-A receptor 8 (EC 2.7.1.112) (tyrosine-protein kinase receptor eek) (eph-and elk-related kinase) (fragment) is designated as the gene product of the gene: *epha8 or eek.* It is a Type I membrane bound receptor, and its function is to serve as a receptor for members of the ephrin-a family. Its catalytic activity is as a protein tyrosine kinase, phosphorylating tyrosine in appropriate target proteins. It is similar to other protein-tyrosine kinases in the catalytic domain and belongs to the ephrin receptor family.

Eph receptors have tyrosine-kinase activity, and, together with their ephrin ligands, mediate contact-dependent cell interactions that are implicated in the repulsion mechanisms that guide migrating cells and neuronal growth cones to specific destinations. Since Eph receptors and ephrins have complementary expression in many tissues during embryogenesis, bidirectional activation of Eph receptors may occur at interfaces of their expression domains, for example, at segment boundaries in the vertebrate hindbrain. Indeed, Eph receptors play key roles in development of the nervous system and angiogenesis. In the nervous system, they

ATCAGAATGGACAGGCACCCCCACCTGTCTTCCTGCCTCTGCATCACCCCCCGGGAAAGCTCCCAGAGCC
CCAGTTCTATGCGGAACCCCACACCTACGAGGAGCCAGGCCGGGCGGGCCGCAGTTTCACTCGGGAGATC
GAGGCCTCTAGGATCCACATCGAGAAAATCATCGGCTCTGGAGACTCCGGGGAAGTCTGCTACGGGAGGC
TGCGGGTGCCAGGGCAGCGGGATGTGCCCGTGGCCATCAAGGCCCTCAAAGCCGGCTACACGGAGAGACA
GAGGCGGGACTTCCTGAGCGAGGCGTCCATCATGGGGCAATTCGACCATCCCAACATCATCCGCCTCGAG
GGTGTCGTCACCCGTGGCCGCCTGGCAATGATTGTGACTGAGTACATGGAGAACGGCTCTCTGGACACCT
TCCTGAGGACCCACGACGGGCAGTTCACCATCATGCAGCTGGTGGGCATGCTGAGAGGAGTGGGTGCCGG
CATGCGCTACCTCTCAGACCTGGGCTATGTCCACCGAGACCTGGCCGCCCGCAACGTCCTGGTTGACAGC
AACCTGGTCTGCAAGGTGTCTGACTTCGGGCTCTCACGGGTGCTGGAGGACGACCCGGATGCTGCCTACA
CCACCACGGGCGGGAAGATCCCCATCCGCTGGACGGCCCCAGAGGCCATCGCCTTCCGCACCTTCTCCTC
GGCCAGCGACGTGTGGAGCTTCGGCGTGGTCATGTGGGAGGTGCTGGCCTATGGGGAGCGGCCCTACTGG
AACATGACCAACCGGGATGTGATCAGCTCTGTGGAGGAGGGGTACCGCCTGCCCGCACCCATGGGCTGCC
CCCACGCCCTGCACCAGCTCATGCTCGACTGTTGGCACAAGGACCGGGCGCAGCGGCCTCGCTTCTCCCA
GATTGTCAGTGTCCTCGATGCGCTCATCCGCAGCCCTGAGAGTCTCAGGGCCACCGCCACAGTCAGCAGG
TGCCCACCCCCTGCCTTCGTCCGGAGCTGCTTTGACCTCCGAGGGGGCAGCGGTGGCGGTGGGGGCCTCA
CCGTGGGGGACTGGCTGGACTCCATCCGCATGGGCCGGTACCGAGACCACTTCGCTGCGGGCGGATACTC
CTCTCTGGGCATGGTGCTACGCATGAACGCCCAGGACGTGCGCGCCCTGGGCATCACCCTCATGGGCCAC
CAGAAGAAGATCCTGGGCAGCATTCAGACCATGCGGGCCCAGCTGACCAGCACCCAGGGGCCCCGCCGGC
ACCTCTGA (SEQ ID NO:4)

A representative amino acid sequence of the ephrin type-A receptor 8-like protein of the invention.

MAPARGRLPPALWVVTAAAAAATCVSAARGEVNLLDTSTIHGDWGWLTYPAHGWDSINEVDESFQPIHTYQVCNVMSPNQNNWL
RTSWVPRDGARRVYAEIKFTLRDCNSMPGVLGTCKETFNLYYLESDRDLGASTQESQFLKIDTIAADESFTGADLGVRRLKLNT
EVRSVGPLSKRGFYLAFQDIGACLAILSLRIYYKKCPAMVRNLAAFSEAVTGADSSSLVEVRGQCVRHSEERDTPKMYCSAEGE
WLVPIGKCVCSAGYEERRDACVACELGFYKSAPGDQLCARCPPHSHSAAPAAQACHCDLSYYRAALDPPSSACTRPPSAPVNLI
SSVNGTSVTLEWAPPLDPGGRSDITYNAVCRRCPWALSRCEACGSGTRFVPQQTSLVQASLLVANLLAHMNYSFWIEAVNGVSD
LSPEPRRAAVVNITTNQAAPSQVVVIRQERAGQTSVSLLWQEPEQPNGIILEYEIKYYEKDKEMQSYSTLKAVTTRATVSGLKP
GTRYVFQVRARTSAGCGRFSQAMEVETGKPRPRYDTRTIVWICLTLITGLVVLLLLICKKRHCGYSKAFQDSDEEKMHYQNGQ
APPPVFLPLHHPPGKLPEPQFYAEPHTYEEPGRAGRSFTREIEASRIHIEKIIGSGDSGEVCYGRLRVPGQRDVPVAIKALKAG
YTERQRRDFLSEASIMGQFDHPNIIRLEGVVTRGRLAMIVTEYMENGSLDTFLRTHDGQFTIMQLVGMLRGVGAGMRYLSDLGY
VHRDLAARNVLVDSNLVCKVSDFGLSRVLEDDPDAAYTTTGGKIPIRWTAPEAIAFRTFSSASDVWSFGVVMWEVLAYGERPYW
NMTNRDVISSVEEGYRLPAPMGCPHALHQLMLDCWHKDRAQRPRFSQIVSVLDALIRSPESLRATATVSRCPPPAFVRSCFDLR
GGSGGGGGLTVGDWLDSIRMGRYRDHFAAGGYSSLGMVLRMNAQDVRALGITLMGHQKKILGSIQTMR (SEQ ID NO:5)

Table 6.
Comparison between a NOV2 polypeptide and mouse ephrin type-A receptor 8 precursor

```
>ref|NP_031965.1| Eph receptor A8
 sp|O09127|EPA8_MOUSE EPHRIN TYPE-A RECEPTOR 8 PRECURSOR (TYROSINE-PROTEIN KINASE
           RECEPTOR EEK) (EPH-AND ELK-RELATED KINASE)
 gb|AAB39218.1| (U72207) Eph-and Elk-related kinase [Mus musculus]
          Length = 1004

 Score = 3036 bits (7128), Expect = 0.0
 Identities = 945/992 (95%), Positives = 964/992 (96%), Gaps = 1/992 (0%)

NOV2:  1   MAPARGRLPPALWVVTAAAAAATCVSAARGEVNLLDTSTIHGDWGWLTYPAHGWDSINEV 60
           ||||| || |||||||||||| ||||| ||||||||||||||||||||||||||||||||
Sbjct: 1   MAPARARLSPALWVVTAAAAA-TCVSAGRGEVNLLDTSTIHGDWGWLTYPAHGWDSINEV 59

NOV2:  61  DESFQPIHTYQVCNVMSPNQNNWLRTSWVPRDGARRVYAEIKFTLRDCNSMPGVLGTCKE 120
           ||||+||||||||||||||||||+|||||||||||||||||||+|||||||||
Sbjct: 60  DESFRPIHTYQVCNVMSPNQNNWLRTNWVPRDGARRVYAEIKFTLRDCNSIPGVLGTCKE 119

NOV2:  121 TFNLYYLESDRDLGASTQESQFLKIDTIAADESFTGADLGVRRLKLNTEVRSVGPLSKRG 180
           ||||+|||||||||||||||||||||||||||||||||||||||||| ||||||||
Sbjct: 120 TFNLHYLESDRDLGASTQESQFLKIDTIAADESFTGADLGVRRLKLNTEVRGVGPLSKRG 179

NOV2:  181 FYLAFQDIGACLAILSLRIYYKKCPAMVRNLAAFSEAVTGADSSSLVEVRGQCVRHSEER 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 180 FYLAFQDIGACLAILSLRIYYKKCPAMVRNLAAFSEAVTGADSSSLVEVRGQCVRHSEER 239
```

```
NOV2 : 61  DESFQPIHTYQVCNVMSPNQNNWLRTSWVPRDGARRVYAEIKFTLRDCNSMPGVLGTCKE 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 61  DESFQPIHTYQVCNVMSPNQNNWLRTSWVPRDGARRVYAEIKFTLRDCNSMPGVLGTCKE 120

NOV2 : 121 TFNLYYLESDRDLGASTQESQFLKIDTIAADESFTGADLGVRRLKLNTEVRSVGPLSKRG 180
           TFNLYYLESDRDLGASTQESQFLKIDTIAADESFTGADLGVRRLKLNTEVRSVGPLSKRG
Sbjct: 121 TFNLYYLESDRDLGASTQESQFLKIDTIAADESFTGADLGVRRLKLNTEVRSVGPLSKRG 180

NOV2 : 181 FYLAFQDIGACLAILSLRIYYKKCPAMVRNLAAFSEAVTGADSSSLVEVRGQCVRHSEER 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 181 FYLAFQDIGACLAILSLRIYYKKCPAMVRNLAAFSEAVTGADSSSLVEVRGQCVRHSEER 240

NOV2 : 241 DTPKMYCSAEGEWLVPIGKCVCSAGYEERRDACVACELGFYKSAPGDQLCARCPPHSHSA 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 241 DTPKMYCSAEGEWLVPIGKCVCSAGYEERRDACVACELGFYKSAPGDQLCARCPPHSHSA 300

NOV2 : 301 APAAQACHCDLSYYRAALDPPSSACTRPPSAPVNLISSVNGTSVTLEWAPPLDPGGRSDI 360
           APAAQACHCDLSYYRAALDPPSSACTRPPSAPVNLISSVNGTSVTLEWAPPLDPGGRSDI
Sbjct: 301 APAAQACHCDLSYYRAALDPPSSACTRPPSAPVNLISSVNGTSVTLEWAPPLDPGGRSDI 360

NOV2 : 361 TYNAVCRRCPWALSRCEACGSGTRFVPQQTSLVQASLLVANLLAHMNYSFWIEAVNGVSD 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 361 TYNAVCRRCPWALSRCEACGSGTRFVPQQTSLVQASLLVANLLAHMNYSFWIEAVNGVSD 420

NOV2 : 421 LSPEPRRAAVVNITTNQAAPSQVVVIRQERAGQTSVSLLWQEPEQPNGIILEYEIKYYEK 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 421 LSPEPRRAAVVNITTNQAAPSQVVVIRQERAGQTSVSLLWQEPEQPNGIILEYEIKYYEK 480

NOV2 : 481 DKEMQSYSTLKAVTTRATVSGLKPGTRYVFQVRARTSAGCGRFSQAMEVETGKPRPRYDT 540
           DKEMQSYSTLKAVTTRATVSGLKPGTRYVFQVRARTSAGCGRFSQAMEVETGKPRPRYDT
Sbjct: 481 DKEMQSYSTLKAVTTRATVSGLKPGTRYVFQVRARTSAGCGRFSQAMEVETGKPRPRYDT 540

NOV2 : 541 RTIVWICLTLITGLVVLLLLLICKKRHCGYSKAFQDSDEEKMHYQNGQAPPPVFLPLHHP 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 541 RTIVWICLTLITGLVVLLLLLICKKRHCGYSKAFQDSDEEKMHYQNGQAPPPVFLPLHHP 600

NOV2 : 601 PGKLPEPQFYAEPHTYEEPGRAGRSFTREIEASRIHIEKIIGSGDSGEVCYGRLRVPGQR 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 601 PGKLPEPQFYAEPHTYEEPGRAGRSFTREIEASRIHIEKIIGSGDSGEVCYGRLRVPGQR 660

NOV2 : 661 DVPVAIKALKAGYTERQRRDFLSEASIMGQFDHPNIIRLEGVVTRGRLAMIVTEYMENGS 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 661 DVPVAIKALKAGYTERQRRDFLSEASIMGQFDHPNIIRLEGVVTRGRLAMIVTEYMENGS 720

NOV2 : 721 LDTFLRTHDGQFTIMQLVGMLRGVGAGMRYLSDLGYVHRDLAARNVLVDSNLVCKVSDFG 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 721 LDTFLRTHDGQFTIMQLVGMLRGVGAGMRYLSDLGYVHRDLAARNVLVDSNLVCKVSDFG 780

NOV2 : 781 LSRVLEDDPDAAYTTTGGKIPIRWTAPEAIAFRTFSSASDVWSFGVVMWEVLAYGERPYW 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 781 LSRVLEDDPDAAYTTTGGKIPIRWTAPEAIAFRTFSSASDVWSFGVVMWEVLAYGERPYW 840

NOV2 : 841 NMTNRDVISSVEEGYRLPAPMGCPHALHQLMLDCWHKDRAQRPRFSQIVSVLDALIRSPE 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 841 NMTNRDVISSVEEGYRLPAPMGCPHALHQLMLDCWHKDRAQRPRFSQIVSVLDALIRSPE 900

NOV2 : 901 SLRATATVSRCPPPAFVRSCFDLRGGSGGGGGLTVGDWLDSIRMGRYRDHFAAGGYSSLG 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 901 SLRATATVSRCPPPAFVRSCFDLRGGSGGGGGLTVGDWLDSIRMGRYRDHFAAGGYSSLG 960

NOV2 : 961 MVLRMNAQDVRALGITLMGHQKKILGSIQTMR 992
           ||||||||||||||||||||||||||||||||
Sbjct: 961 MVLRMNAQDVRALGITLMGHQKKILGSIQTMR 992
```

Table 7.

Multiple alignment of the NOV2 ephrin type-A receptor 8-like protein of the invention, shown as AL035703 Spliced2, with similar proteins

fibronectin like proteins, 99 percent to AAF28459.1 (Lacy et al (1999) Genomics 62 417-426)) and fibromodulin (Table 10, various fibromodulin or fibromodulin-like proteins).

The extracellular matrix (ECM) is composed of collagens, protoglycans, and noncollagenous glycoproteins that provide cells and tissues with a mechanical scaffold for adhesion, migration, and signal transduction (Aumailley and Gayruad (1998) J. Mol. Med. 76(3-4) 253-265). These varied and complex functions depend on interactions between ECM components and cellular receptors such as protoglycans that are located on the cell surface. Fibronectins and fibromodulins are both protoglycans that comprise the extracellular matrix. Disruption of the cell-matrix interactions due to mutations in the genes of the matrix proteins can result in functional failures in all tissues (Bruckner-Tuderman and Bruckner (1998) J. Mol. Med. 76(3-4) 226-237). Included in these disorders are the congenital muscular dystrophies, various muscle disorders, fixed deformities (arthrogryposis), and abnormal white matter by cranial MRI.

Fibronectins are glycoproteins with 2 chains each linked by disulphide bonds that occur in insoluble fibrillar form in the extracellular matrix of animal tissues and soluble in plasma, the latter previously known as cold insoluble globulin. The various slightly different forms of fibronectin appear to be generated by tissue specific differential splicing of fibronectin mRNA, transcribed from a single gene. Fibronectins have multiple domains that confer the ability to interact with many extracellular substances such as collagen, fibrin and heparin and also with specific membrane receptors on responsive cells. Notable is the RGD domain recognized by integrins and two repeats of the EGF like domain. Interaction of a cell's fibronectin receptors (members of the integrin family) with fibronectin adsorbed to a surface results in adhesion and spreading of the cell.

Fibromodulin is collagen-binding protein component of the proteoglycan found in the extracellular matrix. It is mainly expressed in articular cartilage, tendon, and ligament, and is a member of a group of proteins having leucine-rich repeat (LRR) domains; fibromodulin includes as many as ten such motifs. Other components of this family include decorin, biglycan, and lumican. Proteins of this family bind to other matrix macromolecules and thereby help to stabilize the matrix. These proteins may also influence the function of the chondrocytes and bind growth factors.

The core proteins of these proteoglycans are structurally related, consisting of a central region composed of leucine-rich repeats flanked by disulfide-bonded terminal domains. Fibromodulin's central region possesses up to 4 keratan sulfate chains within its leucine-rich domain. Fibromodulin exhibits a wide tissue distribution, with the highest abundance

GGTCATCGACAGCACCACCTGCCCCTCGGTGTGCCGCTGCGACAACGGCTTCATCTACTGCAACGACCGG
GGACTCACATCCATCCCCGCAGATATCCCTGATGATGCCACCACCCTCTACCTGCAGAACAACCAGATCA
ACAACGCCGGCATCCCCCAGGACCTCAAGACCAAGGTCAACGTGCAGGTCATCTACCTATACGAGAATGA
CCTGGATGAGTTCCCCATCAACCTGCCCCGCTCCCTCCGGGAGCTGCACCTGCAGGACAACAATGTGCGC
ACCATTGCCAGGGACTCGCTGGCCCGCATCCCGCTGCTGGAGAAGCTGCACCTGGATGACAACTCCGTGT
CCACCGTCAGCATTGAGGAGGACGCCTTCGCCGACAGCAAACAGCTCAAGCTGCTCTTCCTGAGCCGGAA
CCACCTGAGCAGCATCCCCTCGGGGCTGCCGCACACGCTGGAGGAGCTGCGGCTGGATGACAACCGCATC
TCCACCATCCCGCTGCATGCCTTCAAGGGCCTCAACAGCCTGCGGCGCCTGGTGCTGGACGGTAACCTGC
TGGCCAACCAGCGCATCGCCGACGACACCTTCAGCCGCCTACAGAACCTCACAGAGCTCTCGCTGGTGCG
CAATTCGCTGGCCGCGCCACCCCTCAACCTGCCCAGCGCCCACCTGCAGAAGCTCTACCTGCAGGACAAT
GCCATCAGCCACATCCCCTACAACACGCTGGCCAAGATGCGTGAGCTGGAGCGGCTGGACCTGTCCAACA
ACAACCTGACCACGCTGCCCCGCGGCCTGTTCGACGACCTGGGGAACCTGGCCCAGCTGCTGCTCAGGAA
CAACCCTTGGTTTTGTGGCTGCAACCTCATGTGGCTGCGGGACTGGGTGAAGGCACGGGCGGCCGTGGTC
AACGTGCGGGGCCTCATGTGCCAGGGCCCTGAGAAGGTCCGGGGCATGGCCATCAAGGACATTACCAGCG
AGATGGACGAGTGTTTTGAGACGGGGCCGCAGGGCGGCGTGGCCAATGCGGCTGCCAAGACCACGGCCAG
CAACCACGCCTCTGCCACCACGCCCCAGGGTTCCCTGTTTACCCTCAAGGCCAAAAGGCCAGGGCTGCGC
CTCCCCGACTCCAACATTGACTACCCCATGGCCACGGGTGATGGCGCCAAGACCCTGGCCATCCACGTGA
AGGCCCTGACGGCAGACTCCATCCGCATCACGTGGAAGGCCACGCTCCCCGCCTCCTCTTTCCGGCTCAG
TTGGCTGCGCCTGGGCCACAGCCCAGCCGTGGGCTCCATCACGGAGACCTTGGTGCAGGGGGACAAGACA
GAGTACCTGCTGACAGCCCTGGAGCCCAAGTCCACCTACATCATCTGCATGGTCACCATGGAGACCAGCA
ATGCCTACGTAGCTGATGAGACACCCGTGTGTGCCAAGGCAGAGACAGCCGACAGCTATGGCCCTACCAC
CACACTCAACCAGGAGCAGAACGCTGGCCCCATGGCGAGCCTGCCCCTGGCGGGCATCATCGGCGGGGCA
GTGGCTCTGGTCTTCCTCTTCCTGGTCCTGGGGGCCATCTGCTGGTACGTGCACCAGGCTGGCGAGCTGC
TGACCCGGGAGAGGGCCTACAACCGGGGCAGCAGGAAAAAGGATGACTATATGGAGTCAGGGACCAAGAA
GGATAACTCCATCCTGGAAATCCGCGGCCCTGGGCTGCAGATGCTGCCCATCAACCCGTACCGCGCCAAA
GAGGAGTACGTGGTCCACACTATCTTCCCCTCCAACGGCAGCAGCCTCTGCAAGGCCACACACACCATTG
GCTACGGCACCACGCGGGGCTACCGGGACGGCGGCATCCCCGACATAGACTACTCCTACACA**TGA** (SEQ ID NO:6)

A representative amino acid sequence of the proteoglycan-like protein of the invention

MVVAHPTATATTTPTATVTATVVMTTATMDLRDWLFLCYGLIAFLTEVIDSTTCPSVCRCDNGFIYCNDR
GLTSIPADIPDDATTLYLQNNQINNAGIPQDLKTKVNVQVIYLYENDLDEFPINLPRSLRELHLQDNNVR
TIARDSLARIPLLEKLHLDDNSVSTVSIEEDAFADSKQLKLLFLSRNHLSSIPSGLPHTLEELRLDDNRI
STIPLHAFKGLNSLRRLVLDGNLLANQRIADDTFSRLQNLTELSLVRNSLAAPPLNLPSAHLQKLYLQDN
AISHIPYNTLAKMRELERLDLSNNNLTTLPRGLFDDLGNLAQLLLRNNPWFCGCNLMWLRDWVKARAAVV
NVRGLMCQGPEKVRGMAIKDITSEMDECFETGPQGGVANAAAKTTASNHASATTPQGSLFTLKAKRPGLR
LPDSNIDYPMATGDGAKTLAIHVKALTADSIRITWKATLPASSFRLSWLRLGHSPAVGSITETLVQGDKT
EYLLTALEPKSTYIICMVTMETSNAYVADETPVCAKAETADSYGPTTTLNQEQNAGPMASLPLAGIIGGA
VALVFLFLVLGAICWYVHQAGELLTRERAYNRGSRKKDDYMESGTKKDNSILEIRGPGLQMLPINPYRAK
EEYVVHTIFPSNGSSLCKATHTIGYGTTRGYRDGGIPDIDYSYT (SEQ ID NO:7)

TABLE 9. continued

```
>ref|NP_037412.1| fibronectin leucine rich transmembrane protein 1
 gb|AAF28459.1|AF169675_1 (AF169675) leucine-rich repeat transmembrane protein FLRT1
[Homo
          sapiens]
         Length = 674

 Score = 1365 bits (3494), Expect = 0.0
Identities = 673/674 (99%), Positives = 674/674 (99%)

NOV3:  1   MVVAHPTATATTTPTATVTATVVMTTATMDLRDWLFLCYGLIAFLTEVIDSTTCPSVCRC 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 1   MVVAHPTATATTTPTATVTATVVMTTATMDLRDWLFLCYGLIAFLTEVIDSTTCPSVCRC 60

NOV3:  61  DNGFIYCNDRGLTSIPADIPDDATTLYLQNNQINNAGIPQDLKTKVNVQVIYLYENDLDE 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 61  DNGFIYCNDRGLTSIPADIPDDATTLYLQNNQINNAGIPQDLKTKVNVQVIYLYENDLDE 120

NOV3:  121 FPINLPRSLRELHLQDNNVRTIARDSLARIPLLEKLHLDDNSVSTVSIEEDAFADSKQLK 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 121 FPINLPRSLRELHLQDNNVRTIARDSLARIPLLEKLHLDDNSVSTVSIEEDAFADSKQLK 180

NOV3:  181 LLFLSRNHLSSIPSGLPHTLEELRLDDNRISTIPLHAFKGLNSLRRLVLDGNLLANQRIA 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 181 LLFLSRNHLSSIPSGLPHTLEELRLDDNRISTIPLHAFKGLNSLRRLVLDGNLLANQRIA 240

NOV3:  241 DDTFSRLQNLTELSLVRNSLAAPPLNLPSAHLQKLYLQDNAISHIPYNTLAKMRELERLD 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 241 DDTFSRLQNLTELSLVRNSLAAPPLNLPSAHLQKLYLQDNAISHIPYNTLAKMRELERLD 300

NOV3:  301 LSNNNLTTLPRGLFDDLGNLAQLLLRNNPWFCGCNLMWLRDWVKARAAVVNVRGLMCQGP 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 301 LSNNNLTTLPRGLFDDLGNLAQLLLRNNPWFCGCNLMWLRDWVKARAAVVNVRGLMCQGP 360

NOV3:  361 EKVRGMAIKDITSEMDECFETGPQGGVANAAAKTTASNHASATTPQGSLFTLKAKRPGLR 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 361 EKVRGMAIKDITSEMDECFETGPQGGVANAAAKTTASNHASATTPQGSLFTLKAKRPGLR 420

NOV3:  421 LPDSNIDYPMATGDGAKTLAIHVKALTADSIRITWKATLPASSFRLSWLRLGHSPAVGSI 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 421 LPDSNIDYPMATGDGAKTLAIHVKALTADSIRITWKATLPASSFRLSWLRLGHSPAVGSI 480

NOV3:  481 TETLVQGDKTEYLLTALEPKSTYIICMVTMETSNAYVADETPVCAKAETADSYGPTTTLN 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 481 TETLVQGDKTEYLLTALEPKSTYIICMVTMETSNAYVADETPVCAKAETADSYGPTTTLN 540

NOV3:  541 QEQNAGPMASLPLAGIIGGAVALVFLFLVLGAICWYVHQAGELLTRERAYNRGSRKKDDY 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||+||||
Sbjct: 541 QEQNAGPMASLPLAGIIGGAVALVFLFLVLGAICWYVHQAGELLTRERAYNRGSREKDDY 600

NOV3:  601 MESGTKKDNSILEIRGPGLQMLPINPYRAKEEYVVHTIFPSNGSSLCKATHTIGYGTTRG 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 601 MESGTKKDNSILEIRGPGLQMLPINPYRAKEEYVVHTIFPSNGSSLCKATHTIGYGTTRG 660

NOV3:  661 YRDGGIPDIDYSYT 674
           ||||||||||||||
Sbjct: 661 YRDGGIPDIDYSYT 674
```

This page left blank intentionally.

an N-terminal signal sequence from residue 1 to residue M is cleaved, would have the residues from residue M+1 to residue N remaining. Further as used herein, a "mature" form of a polypeptide or protein may arise from a step of post-translational modification other than a proteolytic cleavage event. Such additional processes include, by way of non-limiting example, glycosylation, myristoylation or phosphorylation. In general, a mature polypeptide or protein may result from the operation of only one of these processes, or a combination of any of them.

The term "probes", as utilized herein, refers to nucleic acid sequences of variable length, preferably between at least about 10 nucleotides (nt), 100 nt, or as many as approximately, e.g., 6,000 nt, depending upon the specific use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are generally obtained from a natural or recombinant source, are highly specific, and much slower to hybridize than shorter-length oligomer probes. Probes can be single- or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

The term "isolated" nucleic acid molecule, as utilized herein, is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5'- and 3'-termini of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated NOV nucleic acid molecules can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb,

0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell/tissue from which the nucleic acid is derived (e.g., brain, heart, liver, spleen, etc.). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material or culture medium when produced by recombinant techniques, or of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecules of the invention, e.g., a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1, 4, or 6, or a complement of this aforementioned nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NO: 1, 4, or 6 as a hybridization probe, NOV molecules can be isolated using standard hybridization and cloning techniques (e.g., as described in Sambrook, et al., (eds.),

compound. Direct binding refers to interactions that do not take place through, or due to, the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

Fragments provided herein are defined as sequences of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, respectively, and are at most some portion less than a full length sequence. Fragments can be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice.

Derivatives are nucleic acid sequences or amino acid sequences formed from the native compounds either directly or by modification or partial substitution. Analogs are nucleic acid sequences or amino acid sequences that have a structure similar to, but not identical to, the native compound but differs from it in respect to certain components or side chains. Analogs can be synthetic or from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type. Homologs are nucleic acid sequences or amino acid sequences of a particular gene that are derived from different species.

Derivatives and analogs can be full length or other than full length, if the derivative or analog contains a modified nucleic acid or amino acid, as described below. Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the invention, in various embodiments, by at least about 30%, 50%, 70%, 80%, or 95% identity (with a preferred identity of 80-95%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the aforementioned proteins under stringent, moderately stringent, or low stringent conditions. See e.g. Ausubel, et al., CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below.

A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences encode those sequences coding for isoforms of NOV polypeptides. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the invention, homologous

cells from a subject e.g., detecting NOV mRNA levels or determining whether a genomic NOV gene has been mutated or deleted.

"A polypeptide having a biologically-active portion of a NOV polypeptide" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a "biologically-active portion of NOV" can be prepared by isolating a portion of SEQ ID NO:1, 4, or 6, that encodes a polypeptide having a NOV biological activity (the biological activities of the NOV proteins are described below), expressing the encoded portion of NOV protein (e.g., by recombinant expression in vitro) and assessing the activity of the encoded portion of NOV.

NOV Nucleic Acid and Polypeptide Variants

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences shown in SEQ ID NO:1, 4, or 6, due to degeneracy of the genetic code and thus encode the same NOV proteins as that encoded by the nucleotide sequences shown in SEQ ID NO:1, 4, or 6. In another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence shown in SEQ ID NO:2, 5, or 7.

In addition to the NOV nucleotide sequences shown in SEQ ID NO:1, 4, or 6, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the NOV polypeptides may exist within a population (e.g., the human population). Such genetic polymorphism in the NOV genes may exist among individuals within a population due to natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame (ORF) encoding a NOV protein, preferably a vertebrate NOV protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the NOV genes. Any and all such nucleotide variations and resulting amino acid polymorphisms in the NOV polypeptides, which are the result of natural allelic variation and that do not alter the functional activity of the NOV polypeptides, are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding NOV proteins from other species, and thus that have a nucleotide sequence that differs from the human sequence of SEQ ID NO:1, 4, or 6, are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the NOV cDNAs of the invention can be isolated based on their homology to the human NOV nucleic acids disclosed herein using the human

65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions are hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C, followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequences of SEQ ID NO:1, 4, or 6, corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

In a second embodiment, a nucleic acid sequence that can hybridize to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, 4, 6, or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X SSC, 5X Denhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55°C, followed by one or more washes in 1X SSC, 0.1% SDS at 37°C. Other conditions of moderate stringency that can be used are well-known within the art. See, e.g., Ausubel ,et al. (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990; GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that can hybridize to the nucleic acid molecule comprising the nucleotide sequences of SEQ ID NO:1, 4, 6, or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that can be used are well known in the art (e.g., as employed for cross-species hybridizations). See, e.g., Ausubel, et al. (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981. Proc Natl Acad Sci USA 78: 6789-6792.

Conservative Mutations

glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted non-essential amino acid residue in the NOV protein is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a NOV coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for NOV biological activity to identify mutants that retain activity. Following mutagenesis of SEQ ID NO:2, 5, or 7, the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

In one embodiment, a mutant NOV protein can be assayed for (i) the ability to form protein:protein interactions with other NOV proteins, other cell-surface proteins, or biologically-active portions thereof, (ii) complex formation between a mutant NOV protein and a NOV ligand; or (iii) the ability of a mutant NOV protein to bind to an intracellular target protein or biologically-active portion thereof; (e.g. avidin proteins). In yet another embodiment, a mutant NOV protein can be assayed for the ability to regulate a specific biological function (e.g., regulation of insulin release).

Antisense Nucleic Acids

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that can hybridize to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:1, 4, 6, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein (e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence). In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire NOV coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a NOV protein of SEQ ID NO:2, 5, or 7; or antisense nucleic acids complementary to a NOV nucleic acid sequence of SEQ ID NO:1, 4, or 6, are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding a NOV protein. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated in situ such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a NOV protein to thereby inhibit expression of the protein (e.g., by inhibiting transcription and/or translation). The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface (e.g., by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens). The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient nucleic acid molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an alpha-anomeric nucleic acid molecule. An alpha-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual alpha-units, the strands run parallel to each other. See, e.g., Gaultier, et al., 1987. Nucl. Acids Res. 15: 6625-6641. The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (see, e.g., Inoue, et al. 1987. Nucl. Acids Res. 15: 6131-6148) or a chimeric RNA-DNA analogue (see, e.g., Inoue, et al., 1987. FEBS Lett. 215: 327-330.

Ribozymes and PNA Moieties

Nucleic acid modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they can be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

In one embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes as described in Haselhoff and Gerlach 1988. Nature 334: 585-591) can be used to catalytically cleave NOV

In another embodiment, PNAs of NOV can be modified, e.g., to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of NOV can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes (e.g., RNase H and DNA polymerases) to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (see, Hyrup, etal., 1996. supra). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup, et al., 1996. supra and Finn, et al., 1996. Nucl Acids Res 24: 3357-3363. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, e.g., 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA. See, e.g., Mag, et al., 1989. Nucl Acid Res 17: 5973-5988. PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment. See, e.g., Finn, et al., 1996. supra. Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. See, e.g., Petersen, et al., 1975. Bioorg. Med. Chem. Lett. 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (e.g., for targeting host cell receptors in vivo), or agents facilitating transport across the cell membrane (see, e.g., Letsinger, et al., 1989. Proc. Natl. Acad. Sci. U.S.A. 86: 6553-6556; Lemaitre, et al., 1987. Proc. Natl. Acad. Sci. 84: 648-652; PCT Publication No. WO88/09810) or the blood-brain barrier (see, e.g., PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (see, e.g., Krol, et al., 1988. BioTechniques 6:958-976) or intercalating agents (see, e.g., Zon, 1988. Pharm. Res. 5: 539-549). To this end, the oligonucleotide can be conjugated to another molecule, e.g., a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, and the like.

NOV Polypeptides

A polypeptide according to the invention includes a polypeptide including the amino acid sequence of NOV polypeptides whose sequences are provided in SEQ ID NO:2, 5, or 7. The invention also includes a mutant or variant protein any of whose residues can be changed from the corresponding residues shown in SEQ ID NO:2, 5, or 7, while still encoding a protein that maintains its NOV activities and physiological functions, or a functional fragment thereof.

NOV chemicals, more preferably less than about 20% chemical precursors or non-NOV chemicals, still more preferably less than about 10% chemical precursors or non-NOV chemicals, and most preferably less than about 5% chemical precursors or non-NOV chemicals.

Biologically-active portions of NOV proteins include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequences of the NOV proteins (e.g., the amino acid sequence shown in SEQ ID NO:2, 5, or 7) that include fewer amino acids than the full-length NOV proteins, and exhibit at least one activity of a NOV protein. Typically, biologically-active portions comprise a domain or motif with at least one activity of the NOV protein. A biologically-active portion of a NOV protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acid residues in length. Moreover, other biologically-active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native NOV protein.

In an embodiment, the NOV protein has an amino acid sequence shown in SEQ ID NO:2, 5, or 7. In other embodiments, the NOV protein is substantially homologous to SEQ ID NO:2, 5, or 7, and retains the functional activity of the protein of SEQ ID NO:2, 5, or 7, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail, below. Accordingly, in another embodiment, the NOV protein is a protein that comprises an amino acid sequence at least about 45% homologous to the amino acid sequence of SEQ ID NO:2, 5, or 7 and retains the functional activity of the NOV proteins of SEQ ID NO:2, 5, or 7.

Determining Homology Between Two or More Sequences

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (i.e., as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology can be determined as the degree of identity between two sequences. The homology can be determined using computer programs known

and the non-NOV polypeptide are fused in-frame with one another. The non-NOV polypeptide can be fused to the N-terminus or C-terminus of the NOV polypeptide.

In one embodiment, the fusion protein is a GST-NOV fusion protein in which the NOV sequences are fused to the C-terminus of the GST (glutathione S-transferase) sequences. Such fusion proteins can facilitate the purification of recombinant NOV polypeptides.
In another embodiment, the fusion protein is a NOV protein containing a heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian host cells), expression and/or secretion of NOV can be increased through use of a heterologous signal sequence.

In yet another embodiment, the fusion protein is a NOV-immunoglobulin fusion protein in which the NOV sequences are fused to sequences derived from a member of the immunoglobulin protein family. The NOV-immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a NOV ligand and a NOV protein on the surface of a cell, to thereby suppress NOV-mediated signal transduction in vivo. The NOV-immunoglobulin fusion proteins can be used to affect the bioavailability of a NOV cognate ligand. Inhibition of the NOV ligand/NOV interaction can be useful therapeutically for both the treatment of proliferative and differentiative disorders, as well as modulating (e.g. promoting or inhibiting) cell survival. Moreover, the NOV-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-NOV antibodies in a subject, to purify NOV ligands, and in screening assays to identify molecules that inhibit the interaction of NOV with a NOV ligand.

A NOV chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, e.g., by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, e.g., Ausubel, et al. (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). A

In addition, libraries of fragments of the NOV protein coding sequences can be used to generate a variegated population of NOV fragments for screening and subsequent selection of variants of a NOV protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a NOV coding sequence with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double-stranded DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, expression libraries can be derived which encodes N-terminal and internal fragments of various sizes of the NOV proteins.

Various techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of NOV proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify NOV variants. See, e.g., Arkin and Yourvan, 1992. Proc. Natl. Acad. Sci. USA 89: 7811-7815; Delgrave, et al., 1993. Protein Engineering 6:327-331.

Anti-NOV Antibodies

The invention encompasses antibodies and antibody fragments, such as Fab or $(Fab)_2$, that bind immunospecifically to any of the NOV polypeptides of said invention.

An isolated NOV protein, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that bind to NOV polypeptides using standard techniques for polyclonal and monoclonal antibody preparation. The full-length NOV proteins can be used or, alternatively, the invention provides antigenic peptide fragments of NOV proteins for use as immunogens. The antigenic NOV peptides comprises at least 4 amino acid residues of the amino acid sequence shown in SEQ ID NO:2, 5, or 7, and encompasses an epitope of NOV such that an antibody raised against the peptide forms a specific immune complex with NOV. Preferably, the antigenic peptide comprises at least 6, 8, 10, 15, 20, or 30 amino acid residues.

The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of NOV. A monoclonal antibody composition thus typically displays a single binding affinity for a particular NOV protein with which it immunoreacts. For preparation of monoclonal antibodies directed towards a particular NOV protein, or derivatives, fragments, analogs or homologs thereof, any technique that provides for the production of antibody molecules by continuous cell line culture can be utilized. Such techniques include, but are not limited to, the hybridoma technique (see, e.g., Kohler & Milstein, 1975. Nature 256: 495-497); the trioma technique; the human B-cell hybridoma technique (see, e.g., Kozbor, et al., 1983. Immunol. Today 4: 72) and the EBV hybridoma technique to produce human monoclonal antibodies (see, e.g., Cole, et al., 1985. In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human monoclonal antibodies can be utilized in the practice of the invention and can be produced by using human hybridomas (see, e.g., Cote, et al., 1983. Proc Natl Acad Sci USA 80: 2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see, e.g., Cole, et al., 1985. In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Each of the above citations is incorporated herein by reference in their entirety.

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to a NOV protein (see, e.g., U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of Fab expression libraries (see, e.g., Huse, et al., 1989. Science 246: 1275-1281) to allow rapid and effective identification of monoclonal Fab fragments with the desired specificity for a NOV protein or derivatives, fragments, analogs or homologs thereof. Non-human antibodies can be "humanized" by techniques well known in the art. See, e.g., U.S. Patent No. 5,225,539. Antibody fragments that contain the idiotypes to a NOV protein can be produced by techniques known in the art including, but not limited to: (i) an $F(ab')_2$ fragment produced by pepsin digestion of an antibody molecule; (ii) an Fab fragment generated by reducing the disulfide bridges of an $F(ab')_2$ fragment; (iii) an Fab fragment generated by the treatment of the antibody molecule with papain and a reducing agent; and (iv) Fv fragments.

Additionally, recombinant anti-NOV antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA

substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, beta-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

NOV Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a NOV protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably-linked" is

Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. Gene 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion E. coli expression vectors include pTrc (Amrann et al., (1988) Gene 69:301-315) and pET 11d (Studier et al., GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

One strategy to maximize recombinant protein expression in E. coli is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein. See, e.g., Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in E. coli (see, e.g., Wada, et al., 1992. Nucl. Acids Res. 20: 2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.
In another embodiment, the NOV expression vector is a yeast expression vector. Examples of vectors for expression in yeast Saccharomyces cerivisae include pYepSec1 (Baldari, et al., 1987. EMBO J. 6: 229-234), pMFa (Kurjan and Herskowitz, 1982. Cell 30: 933-943), pJRY88 (Schultz et al., 1987. Gene 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.). Alternatively, NOV can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., SF9 cells) include the pAc series (Smith, et al., 1983. Mol. Cell. Biol. 3: 2156-2165) and the pVL series (Lucklow and Summers, 1989. Virology 170: 31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987. Nature 329: 840) and pMT2PC (Kaufman, et al., 1987. EMBO J. 6: 187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, and simian virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see, e.g., Chapters 16 and 17 of Sambrook, et al., MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold

not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, NOV protein can be expressed in bacterial cells such as E. coli, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding NOV or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (i.e., express) NOV protein. Accordingly, the invention further provides methods for producing NOV protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding NOV protein has been introduced) in a suitable medium such that NOV protein is produced. In another embodiment, the method further comprises isolating NOV protein from the medium or the host cell.

Spring Harbor, N.Y. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the NOV transgene in its genome and/or expression of NOV mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene-encoding NOV protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a NOV gene into which a deletion, addition or substitution has been introduced to thereby alter, e.g., functionally disrupt, the NOV gene. The NOV gene can be a human gene (e.g., the cDNA of SEQ ID NO:1, 4, or 6), but more preferably, is a non-human homologue of a human NOV gene. For example, a mouse homologue of human NOV gene of SEQ ID NO:1, 4, or 6, can be used to construct a homologous recombination vector suitable for altering an endogenous NOV gene in the mouse genome. In one embodiment, the vector is designed such that, upon homologous recombination, the endogenous NOV gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous NOV gene is mutated or otherwise altered but still encodes functional protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous NOV protein). In the homologous recombination vector, the altered portion of the NOV gene is flanked at its 5'- and 3'-termini by additional nucleic acid of the NOV gene to allow for homologous recombination to occur between the exogenous NOV gene carried by the vector and an endogenous NOV gene in an embryonic stem cell. The additional flanking NOV nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5'- and 3'-termini) are included in the vector. See, e.g., Thomas, et al., 1987. Cell 51: 503 for a description of homologous recombination vectors. The vector is then introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced NOV gene has homologously-recombined with the endogenous NOV gene are selected. See, e.g., Li, et al., 1992. Cell 69: 915.

The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras. See, e.g., Bradley, 1987. In: TERATOCARCINOMAS AND EMBRYONIC STEM CELLS: A PRACTICAL APPROACH, Robertson, ed. IRL, Oxford, pp. 113-152. A chimeric embryo can then be implanted into a suitable pseudopregnant female

antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of Remington's Pharmaceutical Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and non-aqueous vehicles such as fixed oils may also be used. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (i.e., topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid (EDTA); buffers such as acetates, citrates or phosphates, and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (see, e.g., U.S. Patent No. 5,328,470) or by stereotactic injection (see, e.g., Chen, et al., 1994. Proc. Natl. Acad. Sci. USA 91: 3054-3057).

applicable to peptide, non-peptide oligomer or small molecule libraries of compounds. See, e.g., Lam, 1997. Anticancer Drug Design 12: 145.

A "small molecule" as used herein, is meant to refer to a composition that has a molecular weight of less than about 5 kD and most preferably less than about 4 kD. Small molecules can be, e.g., nucleic acids, peptides, polypeptides, peptidomimetics, carbohydrates, lipids or other organic or inorganic molecules. Libraries of chemical and/or biological mixtures, such as fungal, bacterial, or algal extracts, are known in the art and can be screened with any of the assays of the invention.

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt, et al., 1993. Proc. Natl. Acad. Sci. U.S.A. 90: 6909; Erb, et al., 1994. Proc. Natl. Acad. Sci. U.S.A. 91: 11422; Zuckermann, et al., 1994. J. Med. Chem. 37: 2678; Cho, et al., 1993. Science 261: 1303; Carrell, et al., 1994. Angew. Chem. Int. Ed. Engl. 33: 2059; Carell, et al., 1994. Angew. Chem. Int. Ed. Engl. 33: 2061; and Gallop, et al., 1994. J. Med. Chem. 37: 1233.

Libraries of compounds can be presented in solution (e.g., Houghten, 1992. Biotechniques 13: 412-421), or on beads (Lam, 1991. Nature 354: 82-84), on chips (Fodor, 1993. Nature 364: 555-556), bacteria (Ladner, U.S. Patent No. 5,223,409), spores (Ladner, U.S. Patent 5,233,409), plasmids (Cull, et al., 1992. Proc. Natl. Acad. Sci. USA 89: 1865-1869) or on phage (Scott and Smith, 1990. Science 249: 386-390; Devlin, 1990. Science 249: 404-406; Cwirla, et al., 1990. Proc. Natl. Acad. Sci. U.S.A. 87: 6378-6382; Felici, 1991. J. Mol. Biol. 222: 301-310; Ladner, U.S. Patent No. 5,233,409.).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of NOV protein, or a biologically-active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to a NOV protein determined. The cell, for example, can of mammalian origin or a yeast cell. Determining the ability of the test compound to bind to the NOV protein can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the NOV protein or biologically-active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, test compounds can be enzymatically-labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In one embodiment, the assay comprises contacting a cell

operatively linked to a nucleic acid encoding a detectable marker, e.g., luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the invention is a cell-free assay comprising contacting a NOV protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to bind to the NOV protein or biologically-active portion thereof. Binding of the test compound to the NOV protein can be determined either directly or indirectly as described above. In one such embodiment, the assay comprises contacting the NOV protein or biologically-active portion thereof with a known compound which binds NOV to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOV protein, wherein determining the ability of the test compound to interact with a NOV protein comprises determining the ability of the test compound to preferentially bind to NOV or biologically-active portion thereof as compared to the known compound.

In still another embodiment, an assay is a cell-free assay comprising contacting NOV protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to modulate (e.g. stimulate or inhibit) the activity of the NOV protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOV can be accomplished, for example, by determining the ability of the NOV protein to bind to a NOV target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of NOV protein can be accomplished by determining the ability of the NOV protein to further modulate a NOV target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as described, supra.

In yet another embodiment, the cell-free assay comprises contacting the NOV protein or biologically-active portion thereof with a known compound which binds NOV protein to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOV protein, wherein determining the ability of the test compound to interact with a NOV protein comprises determining the ability of the NOV protein to preferentially bind to or modulate the activity of a NOV target molecule.

The cell-free assays of the invention are amenable to use of both the soluble form or the membrane-bound form of NOV protein. In the case of cell-free assays comprising the membrane-bound form of NOV protein, it can be desirable to utilize a solubilizing agent such

conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the NOV protein or target molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity associated with the NOV protein or target molecule.

In another embodiment, modulators of NOV protein expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of NOV mRNA or protein in the cell is determined. The level of expression of NOV mRNA or protein in the presence of the candidate compound is compared to the level of expression of NOV mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of NOV mRNA or protein expression based upon this comparison. For example, when expression of NOV mRNA or protein is greater (i.e., statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of NOV mRNA or protein expression. Alternatively, when expression of NOV mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of NOV mRNA or protein expression. The level of NOV mRNA or protein expression in the cells can be determined by methods described herein for detecting NOV mRNA or protein.

In yet another aspect of the invention, the NOV proteins can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos, et al., 1993. Cell 72: 223-232; Madura, et al., 1993. J. Biol. Chem. 268: 12046-12054; Bartel, et al., 1993. Biotechniques 14: 920-924; Iwabuchi, et al., 1993. Oncogene 8: 1693-1696; and Brent WO 94/10300), to identify other proteins that bind to or interact with NOV ("NOV-binding proteins" or "NOV-bp") and modulate NOV activity. Such NOV-binding proteins are also likely to be involved in the propagation of signals by the NOV proteins as, for example, upstream or downstream elements of the NOV pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for NOV is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, in vivo, forming a NOV-dependent complex, the DNA-binding and activation domains of the

needed enzyme will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. See, e.g., D'Eustachio, et al., 1983. Science 220: 919-924. Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the NOV sequences to design oligonucleotide primers, sub-localization can be achieved with panels of fragments from specific chromosomes. Fluorescence in situ hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical like colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases, will suffice to get good results at a reasonable amount of time. For a review of this technique, see, Verma, et al., HUMAN CHROMOSOMES: A MANUAL OF BASIC TECHNIQUES (Pergamon Press, New York 1988).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, e.g., in McKusick, MENDELIAN INHERITANCE IN MAN, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can then be identified through linkage

greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO:1, 4, or 6, are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

Predictive Medicine

The invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the invention relates to diagnostic assays for determining NOV protein and/or nucleic acid expression as well as NOV activity, in the context of a biological sample (e.g., blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant NOV expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with NOV protein, nucleic acid expression or activity. For example, mutations in a NOV gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with NOV protein, nucleic acid expression, or biological activity.

Another aspect of the invention provides methods for determining NOV protein, nucleic acid expression or activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (e.g., drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (e.g., the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (e.g., drugs, compounds) on the expression or activity of NOV in clinical trials.

These and other agents are described in further detail in the following sections.

Diagnostic Assays

An exemplary method for detecting the presence or absence of NOV in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting NOV protein or nucleic acid (e.g.,

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting NOV protein, mRNA, or genomic DNA, such that the presence of NOV protein, mRNA or genomic DNA is detected in the biological sample, and comparing the presence of NOV protein, mRNA or genomic DNA in the control sample with the presence of NOV protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of NOV in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting NOV protein or mRNA in a biological sample; means for determining the amount of NOV in the sample; and means for comparing the amount of NOV in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect NOV protein or nucleic acid.

Prognostic Assays

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant NOV expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with NOV protein, nucleic acid expression or activity. For example, those involving development, differentiation, and activation of thymic immune cells; in pathologies related to spermatogenesis and male infertility; diagnosis of several human neoplasias; in diseases or pathologies of cells in blood circulation such as red blood cells and platelets; neurological, cardiac and vascular pathologies; rheumatoid arthritis; congenital muscular dystrophies; various muscle disorders; fixed deformities (arthrogryposis); small cell lung cancer NCI-H23; prostate cancer; and abnormal white matter. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disease or disorder. Thus, the invention provides a method for identifying a disease or disorder associated with aberrant NOV expression or activity in which a test sample is obtained from a subject and NOV protein or nucleic acid (e.g., mRNA, genomic DNA) is detected, wherein the presence of NOV protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant NOV expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (e.g., serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein,

sample with one or more primers that specifically hybridize to a NOV gene under conditions such that hybridization and amplification of the NOV gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR can be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (see, Guatelli, et al., 1990. Proc. Natl. Acad. Sci. USA 87: 1874-1878), transcriptional amplification system (see, Kwoh, et al., 1989. Proc. Natl. Acad. Sci. USA 86: 1173-1177); Q Rep licase (see, Lizardi, et al, 1988. BioTechnology 6: 1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a NOV gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, e.g., U.S. Patent No. 5,493,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in NOV can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, to high-density arrays containing hundreds or thousands of oligonucleotides probes. See, e.g., Cronin, et al., 1996. Human Mutation 7: 244-255; Kozal, et al., 1996. Nat. Med. 2: 753-759. For example, genetic mutations in NOV can be identified in two dimensional arrays containing light-generated DNA probes as described in Cronin, et al., supra. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, e.g., U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in NOV genes. For example, single strand conformation polymorphism (SSCP) can be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids. See, e.g., Orita, et al., 1989. Proc. Natl. Acad. Sci. USA: 86: 2766; Cotton, 1993. Mutat. Res. 285: 125-144; Hayashi, 1992. Genet. Anal. Tech. Appl. 9: 73-79. Single-stranded DNA fragments of sample and control NOV nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments can be labeled or detected with labeled probes. The sensitivity of the assay can be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. See, e.g., Keen, et al., 1991. Trends Genet. 7: 5.

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE). See, e.g., Myers, et al., 1985. Nature 313: 495. When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA. See, e.g., Rosenbaum and Reissner, 1987. Biophys. Chem. 265: 12753.

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers can be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions that permit hybridization only if a perfect match is found. See, e.g., Saiki, et al., 1986. Nature 324: 163; Saiki, et al., 1989. Proc. Natl. Acad. Sci. USA 86: 6230. Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See e.g., Eichelbaum, 1996. Clin. Exp. Pharmacol. Physiol., 23: 983-985; Linder, 1997. Clin. Chem., 43: 254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (e.g., N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. At the other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of NOV protein, expression of NOV nucleic acid, or mutation content of NOV genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness

sample from a subject prior to administration of the agent; (ii) detecting the level of expression of a NOV protein, mRNA, or genomic DNA in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the NOV protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the NOV protein, mRNA, or genomic DNA in the pre-administration sample with the NOV protein, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent can be desirable to increase the expression or activity of NOV to higher levels than detected, i.e., to increase the effectiveness of the agent. Alternatively, decreased administration of the agent can be desirable to decrease expression or activity of NOV to lower levels than detected, i.e., to decrease the effectiveness of the agent.

Methods of Treatment

The invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant NOV expression or activity. Such related diseases or disorders include for NOV1 for example, those involving development, differentiation, and activation of thymic immune cells; in pathologies related to spermatogenesis and male infertility; diagnosis of several human neoplasias; in diseases or pathologies of cells in blood circulation such as red blood cells and platelets; and small cell lung cancer NCI-H23; for NOV2, for example, neurological, cardiac and vascular pathologies; for NOV3, for example, rheumatoid arthritis; congenital muscular dystrophies; various muscle disorders; fixed deformities (arthrogryposis); prostate cancer; and abnormal white matter. These methods of treatment will be discussed more fully, below.

Disease and Disorders

Diseases and disorders that are characterized by increased (relative to a subject not suffering from the disease or disorder) levels or biological activity can be treated with Therapeutics that antagonize (i.e., reduce or inhibit) activity. Therapeutics that antagonize activity can be administered in a therapeutic or prophylactic manner. Therapeutics that can be utilized include, but are not limited to: (i) an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; (ii) antibodies to an aforementioned peptide; (iii) nucleic acids encoding an aforementioned peptide; (iv) administration of antisense nucleic acid and nucleic acids that are "dysfunctional" (i.e., due to a heterologous insertion within the coding sequences of coding sequences to an aforementioned peptide) that are utilized to "knockout" endoggenous function of an aforementioned peptide by homologous recombination (see, e.g.,

example, a NOV agonist or NOV antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein. The prophylactic methods of the invention are further discussed in the following subsections.

Therapeutic Methods

Another aspect of the invention pertains to methods of modulating NOV expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of NOV protein activity associated with the cell. An agent that modulates NOV protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of a NOV protein, a peptide, a NOV peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more NOV protein activity. Examples of such stimulatory agents include active NOV protein and a nucleic acid molecule encoding NOV that has been introduced into the cell. In another embodiment, the agent inhibits one or more NOV protein activity. Examples of such inhibitory agents include antisense NOV nucleic acid molecules and anti-NOV antibodies. These modulatory methods can be performed in vitro (e.g., by culturing the cell with the agent) or, alternatively, in vivo (e.g., by administering the agent to a subject). As such, the invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a NOV protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., up-regulates or down-regulates) NOV expression or activity. In another embodiment, the method involves administering a NOV protein or nucleic acid molecule as therapy to compensate for reduced or aberrant NOV expression or activity.

Stimulation of NOV activity is desirable in situations in which NOV is abnormally downregulated and/or in which increased NOV activity is likely to have a beneficial effect. One example of such a situation is where a subject has a disorder characterized by aberrant cell proliferation and/or differentiation (e.g., cancer or immune associated disorders). Another example of such a situation is where the subject has a gestational disease (e.g., preclampsia).

Determination of the Biological Effect of the Therapeutic

In various embodiments of the invention, suitable in vitro or in vivo assays are performed to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected tissue.

In various specific embodiments, in vitro assays can be performed with representative cells of the type(s) involved in the patient's disorder, to determine if a given Therapeutic exerts

The following examples illustrate by way of non-limiting example various aspects of the invention.

Example 1. Quantitative expression analysis of NOV1 and NOV2 in various cells and tissues

The quantitative expression of various clones was assessed in about 41 normal and about 55 tumor samples by real time quantitative PCR (TAQMAN®) performed on a Perkin-Elmer Biosystems ABI PRISM® 7700 Sequence Detection System. In the following Tables 12, 13, and 14, the following abbreviations are used:

ca. = carcinoma,

* = established from metastasis,

met = metastasis,

s cell var= small cell variant,

non-s = non-sm =non-small,

squam = squamous,

pl. eff = pl effusion = pleural effusion,

glio = glioma,

astro = astrocytoma, and

neuro = neuroblastoma

NAT = normal adjacent tissue.

First, up to 96 RNA samples were normalized to β-actin and GAPDH. RNA (~50 ng total or ~1 ng polyA+) was converted to cDNA using the TAQMAN® Reverse Transcription Reagents Kit (PE Biosystems, Foster City, CA; Catalog No. N808-0234) and random hexamers according to the manufacturer's protocol. Reactions were performed in 20 ul and incubated for 30 min. at 48°C. cDNA (5 ul) was then transferred to a separate plate for the TAQMAN® reaction using β-actin and GAPDH TAQMAN® Assay Reagents (PE Biosystems; Catalog Nos. 4310881E and 4310884E, respectively) and TAQMAN® universal PCR Master Mix (PE Biosystems; Catalog No. 4304447) according to the manufacturer's protocol. Reactions were performed in 25 ul using the following parameters: 2 min. at 50°C; 10 min. at 95°C; 15 sec. at 95°C/1 min. at 60°C (40 cycles). Results were recorded as CT values (cycle at which a given sample crosses a threshold level of fluorescence) using a log

A. NOV1

Probe Name: Ag190

Primers	Sequences	Length	Start Position	SEQ ID NO:
Forward	5'-TGGAGGAAGAATCACCACAAGA-3'	22	243	8
Probe	TET-5'-CAAGCCACAAACTGTGACGTGAACCTG-3'-TAMRA	27	271	9
Reverse	5'-GTGGCATCAGCACGGAGTG-3'	19	300	10

The results obtained for clone NOV1 using primer-probe set Ag190 are shown in Table 12.

Table 12.

Tissue_Name	Relative Expression %	Tissue_Name	Relative Expression %
Endothelial cells	0.0	Renal ca. 786-0	0.0
Endothelial cells (treated)	0.0	Renal ca. A498	0.0
Pancreas	0.0	Renal ca. RXF 393	0.0
Pancreatic ca. CAPAN 2	0.0	Renal ca. ACHN	0.9
Adrenal Gland (new lot*)	92.0	Renal ca. UO-31	0.0
Thyroid	0.0	Renal ca. TK-10	0.6
Salavary gland	23.8	Liver	0.0
Pituitary gland	0.0	Liver (fetal)	0.0
Brain (fetal)	0.0	Liver ca. (hepatoblast) HepG2	32.1
Brain (whole)	0.0	Lung	0.0
Brain (amygdala)	0.0	Lung (fetal)	0.0
Brain (cerebellum)	0.0	Lung ca. (small cell) LX-1	0.0
Brain (hippocampus)	0.0	Lung ca. (small cell) NCI-H69	22.9
Brain (thalamus)	0.0	Lung ca. (s.cell var.) SHP-77	0.0
Cerebral Cortex	1.2	Lung ca. (large cell)NCI-H460	4.0
Spinal cord	0.0	Lung ca. (non-sm. cell) A549	8.4
CNS ca. (glio/astro) U87-MG	2.5	Lung ca. (non-s.cell) NCI-H23	100.0
CNS ca. (glio/astro) U-118-MG	0.0	Lung ca (non-s.cell) HOP-62	0.8
CNS ca. (astro) SW1783	0.0	Lung ca. (non-s.cl) NCI-H522	15.4
CNS ca.* (neuro; met) SK-N-AS	7.8	Lung ca. (squam.) SW 900	0.0
CNS ca. (astro) SF-539	0.0	Lung ca. (squam.) NCI-H596	42.3
CNS ca. (astro) SNB-75	0.0	Mammary gland	0.0
CNS ca. (glio) SNB-19	2.1	Breast ca.* (pl. effusion) MCF-7	2.1
CNS ca. (glio) U251	0.0	Breast ca.* (pl.ef) MDA-MB-231	0.0
CNS ca. (glio) SF-295	0.3	Breast ca.* (pl. effusion) T47D	21.0
Heart	0.0	Breast ca. BT-549	0.0
Skeletal Muscle (new lot*)	0.0	Breast ca. MDA-N	16.5
Bone marrow	0.0	Ovary	0.0
Thymus	0.0	Ovarian ca. OVCAR-3	1.0
Spleen	0.0	Ovarian ca. OVCAR-4	0.2
Lymph node	0.0	Ovarian ca. OVCAR-5	1.8
Colorectal	0.0	Ovarian ca. OVCAR-8	7.5
Stomach	0.0	Ovarian ca. IGROV-1	3.9
Small intestine	7.3	Ovarian ca.* (ascites) SK-OV-3	0.5
Colon ca. SW480	0.0	Uterus	0.0
Colon ca.* (SW480 met)SW620	0.4	Plancenta	0.0

Brain (hypothalamus)	0.3	Lung ca. (non-sm. cell) A549	1.1
Spinal cord	2.1	Lung ca. (non-s.cell) NCI-H23	0.6
CNS ca. (glio/astro)U87-MG	0.4	Lung ca (non-s.cell) HOP-62	1.0
CNS ca. (glio/astro)U-118-MG	0.3	Lung ca. (non-s.cl) NCI-H522	0.3
CNS ca. (astro)SW1783	0.3	Lung ca. (squam.) SW 900	11.5
CNS ca.* (neuro; met) SK-N-AS	1.1	Lung ca. (squam.) NCI-H596	0.8
CNS ca. (astro) SF-539	0.0	Mammary gland	1.8
CNS ca. (astro) SNB-75	2.2	Breast ca.* (pl. effusion) MCF-7	0.3
CNS ca. (glio) SNB-19	2.0	Breast ca.* (pl.ef) MDA-MB-231	1.6
CNS ca. (glio) U251	0.9	Breast ca.* (pl. effusion)T47D	0.5
CNS ca. (glio) SF-295	0.0	Breast ca. BT-549	4.7
Heart	0.4	Breast ca. MDA-N	1.6
Skeletal muscle	0.1	Ovary	0.6
Bone marrow	0.1	Ovarian ca. OVCAR-3	0.6
Thymus	3.5	Ovarian ca. OVCAR-4	0.5
Spleen	0.4	Ovarian ca. OVCAR-5	4.6
Lymph node	0.4	Ovarian ca.OVCAR-8	0.3
Colon (ascending)	0.6	Ovarian ca. IGROV-1	0.6
Stomach	1.3	Ovarian ca.* (ascites) SK-OV-3	1.0
Small intestine	0.5	Uterus	1.8
Colon ca. SW480	0.3	Plancenta	1.5
Colon ca.* (SW480 met)SW620	0.2	Prostate	0.5
Colon ca. HT29	2.8	Prostate ca.* (bone met)PC-3	100.0
Colon ca. HCT-116	8.0	Testis	4.6
Colon ca. CaCo-2	1.2	Melanoma Hs688(A).T	0.1
Colon ca. HCT-15	0.9	Melanoma* (met) Hs688(B).T	0.1
Colon ca. HCC-2998	1.5	Melanoma UACC-62	0.8
Gastric ca.* (liver met) NCI-N87	2.8	Melanoma M14	0.3
Bladder	0.4	Melanoma LOX IMVI	0.7
Trachea	1.3	Melanoma* (met)SK-MEL-5	0.2
Kidney	1.7	Melanoma SK-MEL-28	0.3

Normal Kidney GENPAK 061008	0.0	0.0	87071 Bladder Cancer (OD04718-01)	6.0	1.0
83786 Kidney Ca, Nuclear grade 2 (OD04338)	0.0	0.0	87072 Bladder Normal Adjacent (OD04718-03)	2.7	0.0
Tissue_Name/Run_Name	**Relative 2tm723t**	**Expr. % 2tm819t**	**Tissue_Name/Run_Name**	**Relative 2tm723t**	**Expr. % 2tm819t**
83787 Kidney NAT (OD04338)	0.0	0.0	Normal Ovary Res. Gen.	6.5	0.0
83788 Kidney Ca Nuclear grade 1/2 (OD04339)	0.0	0.0	Ovarian Cancer GENPAK 064008	0.0	0.0
83789 Kidney NAT (OD04339)	0.9	0.6	87492 Ovary Cancer (OD04768-07)	25.9	11.7
83790 Kidney Ca, Clear cell type (OD04340)	0.0	0.0	87493 Ovary NAT (OD04768-08)	0.0	0.0
83791 Kidney NAT (OD04340)	0.0	0.0	Normal Stomach GENPAK 061017	0.0	0.0
83792 Kidney Ca, Nuclear grade 3 (OD04348)	0.0	0.0	NAT Stomach Clontech 9060359	0.0	0.0
83793 Kidney NAT (OD04348)	0.0	0.0	Gastric Cancer Clontech 9060395	20.7	27.9
87474 Kidney Cancer (OD04622-01)	0.0	0.0	NAT Stomach Clontech 9060394	1.7	0.0
87475 Kidney NAT (OD04622-03)	0.0	0.0	Gastric Cancer Clontech 9060397	0.0	0.0
85973 Kidney Cancer (OD04450-01)	0.0	0.0	NAT Stomach Clontech 9060396	0.0	0.0
85974 Kidney NAT (OD04450-03)	0.0	0.0	Gastric Cancer GENPAK 064005	0.0	0.0
Kidney Cancer Clontech 8120607	0.0	0.0			

Accordingly the mature form of the predicted extracellular domain of clone NOV2 was targeted for cloning, from residue 28 to 538. Oligonucleotide primers were designed to PCR amplify a DNA segment coding for this mature domain of NOV2. The forward primer includes an in frame BamHI site. The reverse primer contains an in frame XhoI restriction site. The sequences of the primers are the following:

NOV2 Forward: GGATCCGCGCGCGGCGAAGTGAATTTGCTGG (SEQ ID NO:14) and

NOV2 Reverse: CTCGAGGGTCCTGGTGTCATAGCGGGGCC (SEQ ID NO:15).

PCR reactions were set up using 5 ng human hypothalamus cDNA as a template, 1 microM of each of the NOV2 Forward and NOV2 Reverse primers, 5 micromoles dNTP (Clontech Laboratories, Palo Alto CA) and 1 microliter of 50xAdvantage-HF 2 polymerase (Clontech Laboratories, Palo Alto CA) in 50 microliter volume. The following reaction conditions were used:

a) 96°C 3 minutes

b) 96°C 30 seconds denaturation

c) 70°C 30 seconds, primer annealing. This temperature was gradually decreased by 1°C/cycle

d) 72°C 3 minutes extension.

Repeat steps b-d 10 times

e) 96°C 30 seconds denaturation

f) 60°C 30 seconds annealing

g) 72°C 3 minutes extension

Repeat steps e-g 25 times

h) 72°C 5 minutes final extension

A single amplified product having a size of approximately 1500 bp was detected by agarose gel electrophoresis. The product was isolated and ligated into the pCR2.1 vector (Invitrogen Corp, Carlsbad CA).

The construct was sequenced using the following gene-specific primers:

A single amplified product having a size of approximately 1500 bp was detected by agarose gel electrophoresis. The product was isolated and ligated into the pCR2.1 vector (Invitrogen Corp, Carlsbad CA).

The construct was sequenced using the following gene-specific primers:

NOV3S1: CGCACCATTGCCAGGGAC (SEQ ID NO:26),
NOV3S2: GTCCCTGGCAATGGTGCG (SEQ ID NO:27),
NOV3S3: CTGGTGCGCAATTCGCTGGCC (SEQ ID NO:28),
NOV3S4: GGCCAGCGAATTGCGCACCAG (SEQ ID NO:29),
NOV3S5: CACGCCTCTGCCACCACG (SEQ ID NO:30), and
NOV3S6; CGTGGTGGCAGAGGCGTG (SEQ ID NO:31).

The cloned insert was verified as being 100% identical to clone NOV3 (SEQ ID NO:6) from residues 52 to 546. The construct is called pCR2.1-cgNOV3-S331-3A.

Example 5. Preparation of mammalian expression vector pCEP4/Sec

The oligonucleotide primers,

pSec-V5-His Forward: CTCGTCCTCGAGGGTAAGCCTATCCCTAAC
(SEQ ID NO:32) and
pSec-V5-His Reverse: CTCGTCGGGCCCCTGATCAGCGGGTTTAAAC
(SEQ ID NO:33),

were designed to amplify a fragment from the pcDNA3.1-V5His (Invitrogen, Carlsbad, CA) expression vector that includes V5 and His6. The PCR product was digested with XhoI and ApaI and ligated into the XhoI/ApaI digested pSecTag2 B vector harboring an Ig kappa leader sequence (Invitrogen, Carlsbad CA). The correct structure of the resulting vector, pSecV5His, including an in-frame Ig-kappa leader and V5-His6 was verified by DNA sequence analysis. The vector pSecV5His was digested with PmeI and NheI to provide a fragment retaining the above elements in the correct frame. The PmeI-NheI fragment was ligated into the

used was SeeBlue Marker (Invitrogen, Carlsbad, CA). This result is in reasonable agreement with the predicted molecular weight of 54572.3 Da. The program PROSITE predicts that there are three N-glycosylation sites in this polypeptide. Glycosylation of the polypeptide produced in the transfected cells may be responsible for the difference in the molecular weights. The program PROSITE predicts that there are two N-glycosylation sites in this polypeptide. Glycosylation of the polypeptide expressed in the transfected cells may be responsible for the difference between the predicted and observed molecular weights.

Other Embodiments

It is to be understood that while the invention has been described in conjunction with the detailed description thereof, the foregoing description is intended to illustrate and not limit the scope of the invention, which is defined by the scope of the appended claims. Other aspects, advantages, and modifications are within the scope of the following claims.

(b) a variant of a mature form of an amino acid sequence selected from the group consisting of SEQ ID NO: 2, 5, and 7, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 15% of the amino acid residues from the amino acid sequence of said mature form;

(c) an amino acid sequence selected from the group consisting of SEQ ID NO: 2, 5, and 7;

(d) a variant of an amino acid sequence selected from the group consisting of SEQ ID NO: 2, 5, and 7, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 15% of amino acid residues from said amino acid sequence;

(e) a nucleic acid fragment encoding at least a portion of a polypeptide comprising an amino acid sequence chosen from the group consisting of SEQ ID NO: 2, 5, and 7, or a variant of said polypeptide, wherein one or more amino acid residues in said variant differs from the amino acid sequence of said mature form, provided that said variant differs in no more than 15% of amino acid residues from said amino acid sequence; and

(f) a nucleic acid molecule comprising the complement of (a), (b), (c), (d) or (e).

6. The nucleic acid molecule of claim 5, wherein the nucleic acid molecule comprises the nucleotide sequence of a naturally-occurring allelic nucleic acid variant.

7. The nucleic acid molecule of claim 5, wherein the nucleic acid molecule encodes a polypeptide comprising the amino acid sequence of a naturally-occurring polypeptide variant.

8. The nucleic acid molecule of claim 5, wherein the nucleic acid molecule differs by a single nucleotide from a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1, 4, and 6.

9. The nucleic acid molecule of claim 5, wherein said nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of

(a) a nucleotide sequence selected from the group consisting of SEQ ID NO: 1, 4, and 6;

(b) contacting the sample with an antibody that binds immunospecifically to the polypeptide; and

(c) determining the presence or amount of antibody bound to said polypeptide,

thereby determining the presence or amount of polypeptide in said sample.

19. A method for determining the presence or amount of the nucleic acid molecule of claim 5 in a sample, the method comprising:

(a) providing the sample;

(b) contacting the sample with a probe that binds to said nucleic acid molecule; and

(c) determining the presence or amount of the probe bound to said nucleic acid molecule,

thereby determining the presence or amount of the nucleic acid molecule in said sample.

20. The method of claim19 wherein the nucleic acid comprises SEQ ID NO:1 or variants thereof and the sample is chosen from the group consisting of postmeotic hapolid speratids, developing or embryonic tissues, tumor cells, or T cell types.

21. The method of claim 19 wherein the nucleic acid comprises SEQ ID NO:4 or variants thereof and the sample is chosen from the group consisting of cells that express GPI-anchored ephrin-A ligands or mimics thereof.

22. The method of claim 19 wherein the nucleic acid comprises SEQ ID NO:6 or variants thereof and the sample is chosen from the group consisting of tissue from brain or kidney.

23. A method of identifying an agent that binds to a polypeptide of claim 1, the method comprising:

(a) contacting said polypeptide with said agent; and

(b) determining whether said agent binds to said polypeptide.

24. The method of claim 23 wherein the polypeptide comprises SEQ ID NO:2 or variants thereof and the agent is a compound that can modulate actin polymerization or the formation and stability of actin complexes with SEQ ID NO:2 or variants thereof.

34. A pharmaceutical composition comprising the nucleic acid molecule of claim 5 and a pharmaceutically-acceptable carrier.

35. A pharmaceutical composition comprising the antibody of claim 15 and a pharmaceutically-acceptable carrier.

36. A kit comprising in one or more containers, the pharmaceutical composition of claim 33.

37. A kit comprising in one or more containers, the pharmaceutical composition of claim 34.

38. A kit comprising in one or more containers, the pharmaceutical composition of claim 35.

39. The use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a human disease, the disease selected from a NOV-associated disorder, wherein said therapeutic is selected from the group consisting of a NOV polypeptide, a NOV nucleic acid, and a NOV antibody.

40. A method for screening for a modulator of activity or of latency or predisposition to a NOV-associated disorder, said method comprising:

(a) administering a test compound to a test animal at increased risk for a NOV-associated disorder, wherein said test animal recombinantly expresses the polypeptide of claim 1;

(b) measuring the activity of said polypeptide in said test animal after administering the compound of step (a);

(c) comparing the activity of said protein in said test animal with the activity of said polypeptide in a control animal not administered said polypeptide, wherein a change in the activity of said polypeptide in said test animal relative to said control animal indicates the test compound is a modulator of latency of or predisposition to a NOV-associated disorder.

41. The method of claim 40, wherein said test animal is a recombinant test animal that expresses a test protein transgene or expresses said transgene under the control of a promoter at an increased level relative to a wild-type test animal, and wherein said promoter is not the native gene promoter of said transgene.

42. A method for determining the presence of or predisposition to a disease associated with

48. A method of treating a pathological state in a mammal, the method comprising administering to the mammal the antibody of claim 15 in an amount sufficient to alleviate the pathological state.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
26 April 2001 (26.04.2001)

PCT

(10) International Publication Number
WO 01/29217 A3

(51) International Patent Classification[7]: C12N 15/12, 15/16, C07K 14/575, 14/705, 14/47, G01N 33/68, C12Q 1/68, C07K 16/18, 16/26, 16/28, C12N 5/10, A61K 31/7088, 38/17, 38/22, 39/395, 48/00

(21) International Application Number: PCT/US00/28474

(22) International Filing Date: 13 October 2000 (13.10.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/159,805	15 October 1999 (15.10.1999)	US
60/159,992	18 October 1999 (18.10.1999)	US
60/160,952	22 October 1999 (22.10.1999)	US
Not furnished	12 October 2000 (12.10.2000)	US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US	60/159,805 (CIP)
Filed on	15 October 1999 (15.10.1999)
US	60/159,992 (CIP)
Filed on	18 October 1999 (18.10.1999)
US	60/160,952 (CIP)
Filed on	22 October 1999 (22.10.1999)
US	Not furnished (CIP)
Filed on	12 October 2000 (12.10.2000)

(71) Applicant *(for all designated States except US)*: **CURAGEN CORPORATION** [US/US]; 555 Long Wharf Drive, 11th floor, New Haven, CT 06511 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **PRAYAGA, Sudhirdas, K.** [IN/US]; 140 Mill Street, Apt. 18-136, East Haven, CT 06512 (US). **TAUPIER, Raymond, J., Jr.** [US/US]; 407 Holmes Street, East Haven, CT 06512 (US). **BANDARU, Raj** [IN/US]; 3C 107 Cottage Street, New Haven, CT 06511 (US).

(74) Agent: **ELRIFI, Ivor, R.**; Mintz, Levin, Cohn, Ferris, Glovsky, and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

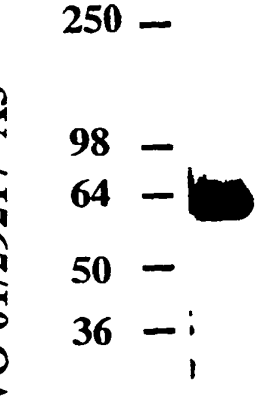
Published:
— *with international search report*

(88) Date of publication of the international search report:
10 January 2002

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES ENCODING SAME

(57) Abstract: Disclosed herein are novel human nucleic acid sequences. Also disclosed are polypeptides encoded by these nucleic acid sequences, and antibodies which immunospecifically-bind to the polypeptide, as well as derivatives, variants, mutants, or fragments of the aforementioned polypeptide, polynucleotide, or antibody. The invention further discloses therapeutic, diagnostic and research methods for diagnosis, treatment, and prevention of disorders involving these novel nucleic acids and proteins.



WO 01/29217 A3

INTERNATIONAL SEARCH REPORT

I. ıational Application No
PCT/US 00/28474

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 A61K48/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 99 47674 A (CORIXA CORP) 23 September 1999 (1999-09-23) SEQ ID NO:83 page 1 -page 32; claims 1-58; examples 1-4 --- -/--	5-14,19, 20,29, 30,34, 37,39, 43-45,47

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
20 June 2001	

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016	Devijver, K

Form PCT/ISA/210 (second sheet) (July 1992)

5

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 00/28474

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	DATABASE EMBL 'Online! accession: AF169675, 31 January 2000 (2000-01-31) LACY S E ET AL: "Homo sapiens leucine-rich repeat transmembrane protein FLRT1 (FLRT1) mRNA, complete cds." XP002169956 -& LACY SUSAN E ET AL: "Identification of FLRT1, FLRT2, and FLRT3: A novel family of transmembrane leucine-rich repeat proteins." GENOMICS, vol. 62, no. 3, 15 December 1999 (1999-12-15), pages 417-426, XP002169951	1-17, 33-38
P,X	DATABASE EMBL 'Online! accession: AP000597, 15 October 1999 (1999-10-15) HATTORI M ET AL: "Homo sapiens genomic DNA, chromosome 11q13 clone:RP11-45M18, WORKING DRAFT SEQUENCE, 14 unordered pieces. " XP002169957	5-14,34, 37
A	WETERMAN M A J ET AL: "THYMOSIN BETA-10 EXPRESSION IN MELANOMA CELL LINES AND MELANOCYTIC LESIONS: A NEW PROGRESSION MARKER FOR HUMAN CUTANEOUS MELANOMA" INTERNATIONAL JOURNAL OF CANCER,US,NEW YORK, NY, vol. 53, no. 2, 21 January 1993 (1993-01-21), pages 278-284, XP002041675 ISSN: 0020-7136	
A	TUERECI OE ET AL: "HUMAN CARBONIC ANHYDRASE XII: CDNA CLONING,EXPRESSION, AND CHROMOSOMAL LOCALIZATION OF A CARBONIC ANHYDRASE GENE THAT IS OVEREXPRESSED IN SOME RENAL CELL CANCERS" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA,NATIONAL ACADEMY OF SCIENCE. WASHINGTON,US, vol. 95, no. 13, 23 June 1998 (1998-06-23), pages 7608-7613, XP000941354 ISSN: 0027-8424	

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

Invention 1:
Claims: in part 1-19,23,25-44,47,48 (all as far as applicable); completely: 20,24,45

Polynucleotide and polypeptide relating to SEQ ID NOs 1 and 2, and variants and fragments thereof. Expression vector and host cells comprising such a polynucleotide. Antibody specifically binding with such a polypeptide. Method for detecting such a polynucleotide or such a polypeptide. Method for identifying an agent that binds to such a polypeptide or modulates the expression or activity of such a polypeptide. Method for treating or preventing a disorder by using such a polynucleotide, polypeptide or antibody. Pharmaceutical composition comprising such a polynucleotide, polypeptide or antibody. Kit comprising such a composition. Use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a disease.

Invention 2:
Claims: in part 1-19,23,25-44,47,48 (all as far as applicable); completely: 21,46

As invention 1, but limited to subject-matter relating to SEQ ID NOs 4 and 5.

Invention 3
Claims: in part 1-19,23,25-44,47,48 (all as far as applicable); completely: 22

As invention 1, but limited to subject-matter relating to SEQ ID NOs 6 and 7.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 00/28474

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9947674 A	23-09-1999	US 6210883 B	03-04-2001
		AU 3094999 A	11-10-1999
		BR 9908823 A	21-11-2000
		EP 1064372 A	03-01-2001
		NO 20004631 A	15-11-2000